Ann Inst Stat Math (2012) 64:275–301
DOI 10.1007/s10463-010-0311-8

Modelling time trend via spline confidence band

Jing Wang

Received: 4 January 2008 / Revised: 28 April 2010 / Published online: 31 July 2010

Abstract Simultaneous confidence band is obtained for the trend function of time series with heteroscedastic α-mixing errors, based on constant and linear spline smoothing. Simulation study confirms that the bands have conservative coverage of the true trend function. Linear band has been constructed for the leaf area index (LAI) data collected in East Africa, which has revealed that the trigonometric curve in the regional atmospheric modelling system (RAMS) is inadequate.

Keywords Berry–Esseen bound · Confidence band · Heteroscedastic error · Mixing · Polynomial spline · Trend

1 Introduction

In time series analysis, estimation of the trend is a very crucial first step. Consider a time series realization $\{Y_i\}_{i=1}^n$, one common assumption is that Y_i is decomposable into a time varying trend plus a stationary error, i.e., $Y_i = m(i/n) + X_i$ in which $m(\cdot)$ is the trend function and the errors $\{X_i\}_{i=1}^n$ form a time series with mean zero. Often it is assumed that the trend function m is of polynomial form, and the errors $\{X_i\}_{i=1}^n$ is stationary. Many actual time series data, however, exhibit strong deviation from these set of assumptions, see for example, Beran and Feng (2002a), Beran and Feng (2002b), and Feng (2004).

J. Wang (✉)
Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA
e-mail: wangjing@math.uic.edu

Without loss of generality, consider the following model that incorporates nonparametric time trend and heteroscedastic error

$$Y_{in} = m\,(i/n) + \sigma\,(i/n)\,\varepsilon_{in}, \quad i = 1, \ldots, n \tag{1}$$

in which $m(i/n)$ is a smooth function of time i/n, the error X_i is expressed as $\sigma(i/n)\varepsilon_{in}$, where $\sigma(i/n)$ is the standard deviation of X_i and $\{\varepsilon_{in}\}_{i=1}^{n}$ is a standardized white noise sequence, $E(\varepsilon_{in}) \equiv 0$, $E(\varepsilon_{in}^2) \equiv 1$.

Nonparametric regression problems have been long investigated through kernel and spline smoothing in parallel. For kernel smoothing on independent data, satisfactory asymptotics have been obtained for pointwise convergence, see Fan and Gijbels (1996) and Härdle (1990). The uniform confidence bands of the kernel type are also available in Bickel and Rosenblatt (1973), Silverman (1986), Härdle (1989), Xia (1998) and Claeskens and Van Keilegom (2003). Most of the technical part for uniform convergence rely heavily on the strong approximation theorem in Tusnády (1977). Practitioners especially favor superior computational performance and easy implementation procedure. Polynomial spline smoothers perform satisfactorily in both aspects. The detailed asymptotics has been developed in Stone (1985), Stone (1994), Huang (1998) and Huang (2003). Uniform confidence bands are also available in Zhou et al. (1998) under homoscedastic, independent and normality assumption, and Wang and Yang (2009a) under independent restriction only. Wang and Yang (2009b) has constructed kernel type uniform confidence bands for component function in high-dimensional additive models.

Nonparametric smoothing of weakly dependent data has been pursued in many directions due to its superiority for modelling and forecasting nonlinear time series, see Roussas (1988) and Roussas (1990) for kernel type smoothing, Fan and Yao (2003) for local polynomial fitting, Huang and Yang (2004), and Xue and Yang (2006) for spline type smoothing, and Cai (2002) for regression quantiles. Song and Yang (2009) investigate the asymptotics normality of variance spline estimator. Under strong mixing conditions asymptotic normality for pointwise estimators are available in the cases of kernel regression in Liebscher (1999), and polynomial spline regression in Diack (2001) and Huber-Carol et al. (2002). In addition Masry and Fan (1997) established asymptotic joint normal distribution for derivatives of regression function. Rates for uniform (strong) convergence under mixing condition was proposed for Nadaraya–Watson regression estimation in Liebscher (2001). Zhou et al. (1998) constructed the simultaneous confidence band for polynomial splines under i.i.d. cases. However, uniform confidence bands for dependent observations, in particular in the strong mixing condition, are unavailable for the nonparametric smoothers.

As an example, consider regional climate modelling in Olson et al. (2008), of the phenological information, such as Leaf Area Index (LAI) and fractional cover, which are important inputs of the climate modelling system. With modern satellite system, the data are regularly collected over recent years. The observation does show clear pattern of heteroscedasticity and dependence over time. It can be best described by model (1). One major objective of the system is to update LAI in the model system if the overall shape underlying the real data does not support the original LAI function.

Such inference about the regression function is best done with the construction of uniform confidence bands for hypothesized function.

The paper is organized as follows. Section 2 introduces main results about the confidence bands on time series data. Implementation description and simulation results are presented in Sects. 3 and 4. Data example illustrates the proposed model in Sect. 5. Outline of the proof is in Appendix.

2 Main results

To introduce spline functions, divide interval [0, 1] into $(N+1)$ subintervals $J_j = [t_j, t_{j+1})$, $j = 0, \dots, N-1$, $J_N = [t_N, 1]$. A sequence of equally spaced points $\{t_j\}_{j=1}^N$, called interior knots, are given as

$$t_{1-p} = \cdots = t_0 = 0 < t_1 < \cdots < t_N < 1 = t_{N+1}, t_j = a + jh, \quad j = 0, 1, \dots, N+1, \tag{2}$$

in which $h = 1/(N+1)$ is the distance between neighboring knots and p is to denote spline order. Order $p = 1, 2$ represent constant and linear spline respectively. We denote by $G^{(p-2)} = G^{(p-2)}[0, 1]$ the space of functions that are polynomials of degree $(p-1)$ on each J_j and have continuous $(p-2)$th derivative on [0, 1]. For example, $G^{(-1)}$ denotes the space of functions that are constant on each J_j, and $G^{(0)}$ the space of functions that are linear on each J_j and continuous on [0, 1]. Denote by $C^{(p)}[0, 1]$ the function space in which every function has continuous pth order derivative.

Define the polynomial spline estimator based on data $\{Y_{in}\}_{i=1}^n$ drawn from model (1)

$$\hat{m}_p(\cdot) = \underset{g(\cdot) \in G^{(p-2)}}{\operatorname{argmin}} \sum_{i=1}^n \{Y_{in} - g(i/n)\}^2, \quad p = 1, 2. \tag{3}$$

Necessary technical assumptions are as follows:

(A1) *The regression function* $m \in C^{(p)}[0, 1]$, $p = 1, 2$.

(A2) *The standard deviation function* $\sigma(x)$ *is continuous and positive on* [0, 1].

(A3) *The errors* $\{\varepsilon_{in}\}_{i=1}^n$ *is a martingale difference that satisfies*

$$E\left(\varepsilon_{in} \left| \mathcal{F}_{i-1,n}\right.\right) = 0, \; E\left(\varepsilon_{in}^2\right) = 1, E\left(|\varepsilon_{in}|^3\right) < M_0 < +\infty, 1 \le i \le n,$$

in which $\mathcal{F}_{i-1,n}$ *denotes the* σ *-field generated by* $\{\varepsilon_{1,n}, \dots, \varepsilon_{i-1,n}\}$.

(A4) *There exist positive constants* K_0 *and* λ_0 *such that* $\alpha_n(k) \le K_0 e^{-\lambda_0 k}$ *holds for all* $1 \le k \le n-1$, *where the strong mixing coefficient of order* k *is defined as*

$$\alpha_n(k) = \sup_{B \in \sigma\{\varepsilon_{in}, i \le t\}, C \in \sigma\{\varepsilon_{in}, i \ge t+k\}} |P(B \cap C) - P(B)P(C)|. \tag{4}$$

(A5) *The number of interior knots* $N = N_n$ *satisfies* $(n/\log n)^{1/(2p+1)} \ll N \ll n^{1/3}$.

Assumptions (A1) and (A2) are the regular assumptions in nonparametric regression, same as in Huang (2003). Assumptions (A3) and (A4) are common ones in time series literature. The knots number N follows order constraints consistent with Stone (1985, 1994).

Define for any $x \in [0, 1]$ its location index $j(x)$ and relative location index $\delta(x)$ as

$$j(x) = j_n(x) = \min\{[x/h], N\}, \ \delta(x) = \left\{x - t_{j(x)}\right\}/h. \tag{5}$$

It is clear that $t_{j(x)} \le x < t_{j(x)+1}, 0 \le \delta(x) < 1, \forall x \in [0, 1], j(1) = N$. Also note that the relative location index equals 0 at all interior knots, i.e. $\delta(t_j) = 0$, for $j = 0, 1, \ldots, N$, and $\delta(0) = 0, \delta(1) = 1$.

For any L^2-integrable functions ϕ, φ on [0, 1], the theoretical and empirical inner products are defined respectively as $\langle \phi, \varphi \rangle = \int_0^1 \phi(x)\varphi(x)\mathrm{d}x$, $\langle \phi, \varphi \rangle_n = n^{-1} \sum_{i=1}^{n} \{\phi(i/n)\varphi(i/n)\}$. The corresponding L^2 norms can be defined accordingly.

The B-spline basis functions that generate $G^{(-1)}$, are indicator functions on subintervals J_j, and defined as $b_{j,1}(x) = I_j(x) = I_{J_j}(x)$, $j = 0, 1, \ldots, N$. The linear B-spline basis of $G^{(0)}$, are $b_{j,2}(x) = K\{(x - t_{j+1})h^{-1}\}$, $j = -1, 0, 1, \ldots, N$, where $K(u) = (1 - |u|)_+$ is the triangular kernel. For theoretical analysis, we use rescaled (standardized) B-spline basis $\{B_{j,1}(x)\}_{j=0}^{N}$ and $\{B_{j,2}(x)\}_{j=-1}^{N}$ for $G^{(-1)}$ and $G^{(0)}$,

$$B_{j,p}(x) \equiv \left\|b_{j,p}\right\|_2^{-1} b_{j,p}(x), 1 - p \le j \le N, p = 1, 2. \tag{6}$$

The inner product matrix of constant B-spline basis $\{B_{j,1}(x)\}_{j=0}^{N}$ is obviously an identity matrix $\mathbf{I}_{N+1}$, while the inner product of linear B-spline basis $\{B_{j,2}(x)\}_{j=-1}^{N}$ is denoted as

$$\mathbf{V} = \left(\langle B_{j',2}, B_{j,2}\rangle\right)_{j,j'=-1}^{N} = \left(v_{j'j}\right)_{j,j'=-1}^{N}. \tag{7}$$

Note that the linear spline basis function only overlaps with its neighbors, matrix $\mathbf{V}$ equals to a tridiagonal matrix which is approximated by (29) in Lemma 7. Define inverse matrix of $\mathbf{V}$ as $\mathbf{S}$, and the 2×2 diagonal submatrices of matrix $\mathbf{S}$ as follows

$$\mathbf{S} = \left(s_{j'j}\right)_{j,j'=-1}^{N} = \mathbf{V}^{-1}, \quad \mathbf{S}_j = \begin{pmatrix} s_{j-1,j-1} & s_{j-1,j} \\ s_{j,j-1} & s_{j,j} \end{pmatrix}, \quad j = 0, \ldots, N. \tag{8}$$

Define

$$\Sigma = \left(\sigma_{jl}^2\right)_{j,l=-1}^{N} = \left\{\int \sigma^2(v) B_{j,2}(v) B_{l,2}(v)\,\mathrm{d}v\right\}_{j,l=-1}^{N}, \tag{9}$$

where the component σ_{jl}^2 is a key quantity for heteroscedastic variance function, which determines the width of confidence bands. The approximate pointwise variance function of the polynomial spline estimator $\hat{m}_p(x)$ for $p = 1, 2$, are

$$\sigma_{n,1}^2(x) = \frac{\int I_{j(x)}(v)\,\sigma^2(v)\,\mathrm{d}v}{n\left\|b_{j(x),1}\right\|_2^2} = \frac{\int I_{j(x)}(v)\,\sigma^2(v)\,\mathrm{d}v}{nh}, \tag{10}$$

$$\sigma_{n,2}^2(x) = \frac{1}{n}\sum_{j,j',l,l'=-1}^{N} B_{j',2}(x)\,B_{l',2}(x)\,s_{jj'}s_{ll'}\sigma_{jl}^2, \tag{11}$$

with $j(x)$ defined in (5), $B_{j',2}(x)$ in (6), and $s_{ll'}$ and σ_{jl} in (8), (9).

The main theorem is to construct simultaneous confidence bands with polynomial spline smoother. An asymptotic conservative $100(1-\alpha)\%$ confidence band for the unknown $m(x)$ over the interval [0, 1] consists of an estimator $\hat{m}(x)$ of $m(x)$, lower and upper confidence limits $\hat{m}(x) - l_n(x)$, $\hat{m}(x) + l_n(x)$ at every $x \in [a, b]$ such that

$$\liminf_{n\to\infty} P\left\{m(x) \in \hat{m}(x) \pm l_n(x),\ \forall x \in [0, 1]\right\} \ge 1-\alpha.$$

Theorem 1 *Under Assumptions* (A1)–(A5)*, for a given* $0 < \alpha < 1$*, an asymptotic* $100(1-\alpha)\%$ *conservative confidence band for* $m(x)$ *over* $[0, 1]$ *is*

$$\hat{m}_p(x) \pm \sigma_{n,p}(x)\,\{2p\log(N+1)\}^{1/2}\,d_n(\alpha/p), \quad p = 1, 2$$

where the pointwise variance $\sigma_{n,1}(x)$ *as given in* (10) *can be approximated by* $\sigma(x)$ $\{nh\}^{-1}$ *according to Lemma* 3*,* $\sigma_{n,2}(x)$ *in* (11) *could be replaceable by* $\sigma(x)\{2nh/3\}^{-1}$ $\mathbf{\Delta}^T(x)\mathbf{S}_{j(x)}\mathbf{\Delta}(x)$ *according to Lemma* 8*, and where*

$$d_n(\alpha) = 1 - \{2\log(N+1)\}^{-1}\left[\log(\alpha/2) + \frac{1}{2}\{\log\log(N+1) + \log 4\pi\}\right]. \tag{12}$$

To develop the asymptotics of the estimator $\hat{m}_p(x)$, estimation error $\hat{m}_p(x) - m(x)$ is decomposed into a bias term and a noise term

The spline estimator $\hat{m}_p(x)$ can be expressed as a linear combination of the standardized B-spline basis, $\hat{m}_p(x) \equiv \sum_{j=1-p}^{N} \hat{\lambda}_{j,p} B_{j,p}(x)$, where the coefficients $\{\hat{\lambda}_{j,p}, 1-p \le j \le N\}^T$ are solutions of the following least squares problem

$$\left\{\hat{\lambda}_{1-p,p}, \ldots, \hat{\lambda}_{N,p}\right\}^T = \operatorname*{argmin}_{R^{N+p}} \sum\nolimits_{i=1}^{n} \left\{Y_i - \sum\nolimits_{j=1-p}^{N} \lambda_{j,p} B_{j,p}(i/n)\right\}^2. \tag{13}$$

Denote a pseudo function $\mathbf{Y}(i/n) \equiv Y_{in}$, $1 \le i \le n$, then $\hat{m}_p(x)$ can be expressed as

$$\hat{m}_p(x) = \left\{B_{j,p}(x)\right\}_{1-p\le j\le N}^T \left(\langle B_{j',p}, B_{j,p}\rangle_n\right)_{1-p\le j,j'\le N}^{-1} \left\{\langle \mathbf{Y}, B_{j,p}\rangle_n\right\}_{j=1-p}^{n},$$

which is the projection of response $\mathbf{Y}$ in spline space $G^{(1-p)}$ with respect to the empirical inner product. Correspondingly we could define projections of true function $\mathbf{m}(x)$ and pseudo function based on error $\mathbf{E}(i/n) \equiv \{\sigma(i/n)\varepsilon_{in}\}_{i=1}^{n}$, $\tilde{m}_p(x)$ and $\tilde{\varepsilon}_p(x)$. The error decomposition can be formulated as

$$\hat{m}_p(x) - m(x) = \left\{\tilde{m}_p(x) - m(x)\right\} + \tilde{\varepsilon}_p(x).$$

For notation simplicity, denote by $\|\cdot\|_\infty$ the supremum norm of a function r on $[0, 1]$, i.e. $\|r\|_\infty = \sup_{x\in[0,1]} |r(x)|$, and the moduli of continuity of a continuous function r on $[0, 1]$ is denoted as $\omega(r, h) = \max_{x,x'\in[0,1],|x-x'|\le h} |r(x) - r(x')|$. By the uniform continuity of r on an interval $[0, 1]$, one has $\lim_{h\to 0}\omega(r, h) = 0$.

The magnitude order of bias term $\tilde{m}_p(x) - m(x)$ can be derived from the theorem on page 149, de Boor (2001), and Theorem 5.1, Huang (2003).

The following proposition provides the asymptotic magnitude of the noise term $\tilde{\varepsilon}_p(x)$, which plays a major role in the proof for Theorem 1 in Appendix.

Proposition 1 *Under Assumptions* (A2)–(A5)*, for a given* $0 < \alpha < 1$, *with* $\sigma_{n,1}(x)$ *given in* (10) *and* $\sigma_{n,2}(x)$ *in* (11)

$$\liminf_{n\to\infty} P\left[\sup_{x\in[0,1]} \left|\sigma_{n,p}^{-1}(x)\,\tilde{\varepsilon}_p(x)\right| \le \{2p\log(N+1)\}^{1/2} d_n(\alpha/p)\right] \ge 1-\alpha.$$

Remark 1 The main theorem results show that the strong convergence rate of strong mixing sequence is similar to that of i.i.d. cases in Wang and Yang (2009a), which is quite intuitive and consistent with other smoothing estimators in both cases. But they are very different particularly in the technical part. In Wang and Yang (2009a) the approximation procedure relies on the strong approximation of empirical processes of i.i.d. observations by Brownian bridge in both constant and linear spline cases. With the Itô Isometry theorem, an exact distribution is available for contact case and hence an exact confidence band; while for linear splines due to the tridiagonal structure of the inner product of the basis vector, a maximization theorem is applied and a conservative band is for linear splines. But in this paper, because of the dependent structure (α-mixing) among the observations, an exact approximation to the empirical processes is not available any more, but a conservative band can be obtained via Berry–Esseen Bound in Sunklodas (1984). Liang and Uña-Álvarez (2009) used similar idea to investigate the uniform asymptotics for censored data under strong mixing conditions. As in Liebscher (2001), Bernstein's inequality in Bosq (1996) and upcrossing probability in Leadbetter et al. (1983) are also popular tools in this paper to develop simultaneous confidence band.

Remark 2 Following the suggestion of the referees and the associate editor, the standard deviation $\sigma(x)$ in Theorem 1 is estimated by its polynomial spline estimator, instead of the kernel type counterpart in implementation of the following section. Song and Yang (2009) has shown that in their simulation studies bootstrap sampling helps to achieve much more powerful regressors. Similar bootstrap methods can be found in Härdle et al. (2004) and Yang (2008) for component function estimation in generalized additive models and additive models respectively in the setting of independent and identically distributed observations. The asymptotic behavior of kernel estimators after bootstrap sampling has been studied in Paparoditis and Politis (2000) under mixing conditions. The author suspects the convergence rate and asymptotic normality of Bootstrap sampling can be extended to other nonparametric smoothing estimators including spline smoothing under similar dependent restrictions. It is an interesting topic which needs further investigation, but beyond the scope of this paper.

3 Implementation

To implement the proposed confidence bands in Theorem 1, a sample $\{Y_{in}\}_{i=1}^{n}$ is drawn from model (1). The number of interior knots for constant spline is taken to be $N = [c_1 n^{1/3} (\log n)^{-1/6}] + c_2$, where c_1 and c_2 are positive integers, and knots number corresponding to linear spline is $N = [c_1 n^{1/5}] + c_2$. The knots are taken to be equally spaced, as in (2). Since explicit formula for coverage probability does not exist for the bands, there is no optimal method to select (c_1, c_2) supported by theoretical inference. Härdle et al. (1997) had shown that adaptive knots selection could lead to inconsistency in L_∞ norm. In simulation, the simple choice of $c_1 = 5$ and $c_2 = 1$ for constant band and $c_1 = 1$ and $c_2 = 1$ for linear band seem work well, so these are set as default values.

The least squares problem for B-spline estimator $\hat{m}_p(\cdot)$ in (13) can be solved via the truncated power basis $\{1, x, \ldots, x^{p-1}, (x - t_j)_+^{p-1}, j = 1, \ldots, N\}$. In the spline space, given knots location and number, the truncated power basis is equivalent to the previous mentioned B-spline basis functions in (6), i.e.

$$\hat{m}_p(x) = \sum_{k=0}^{p-1} \hat{\gamma}_k x^k + \sum_{j=1}^{N} \hat{\gamma}_{j,p} \left(x - t_j\right)_+^{p-1}, \quad p = 1, 2, \tag{14}$$

where the coefficients $\{\hat{\gamma}_0, \ldots, \hat{\gamma}_{p-1}, \hat{\gamma}_{1,p}, \ldots, \hat{\gamma}_{N,p}\}^T$ are solutions to the following least squares problem

$$\left\{\hat{\gamma}_0, \ldots, \hat{\gamma}_{N,p}\right\}^T = \underset{R^{N+p}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{ Y_i - \sum_{k=0}^{p-1} \gamma_k (i/n)^k - \sum_{j=1}^{N} \gamma_{j,p} \left((i/n) - t_j\right)_+^{p-1} \right\}^2 .$$

Define vectors $\mathbf{Z}_p = \{Z_{1,p}, \ldots, Z_{n,p}\}^T$, $p = 1, 2$, with $Z_{i,p} = \{Y_i - \hat{m}_p(i/n)\}^2$. Following two referees' suggestion, the polynomial spline estimation is applied to obtain the variance estimator of $\sigma^2(x)$, $\hat{\sigma}_p^2(x)$, $p = 1, 2$, based on data $\{i/n, Z_{i,p}\}_{i=1}^{n}$

$$\hat{\sigma}_p^2(x) = \underset{g(\cdot) \in G^{(p-2)}}{\operatorname{argmin}} \sum_{i=1}^{n} \left\{Z_{i,p} - g(i/n)\right\}^2, \quad p = 1, 2. \tag{15}$$

The consistency of the estimator is obtained in Song and Yang (2009)

$$\max_{p=1,2} \sup_{x\in[a,b]} \left|\hat{\sigma}_p^2(x) - \sigma^2(x)\right| = O_p\left(n^{-p/(2p+1)}\right) = o_p(1). \tag{16}$$

The conservative band is computed with $\hat{m}_p(x)$ in (14) and d_n in (28)

$$\hat{m}_p(x) \pm \hat{\sigma}_{n,p}(x)\{2p\log(N+1)\}^{1/2} d_n(\alpha/p) \tag{17}$$

where the function $\sigma_{n,p}(x)$ is approximated by two formulae given below. When $p=1$, $\sigma_{n,1}(x)$ is approximated by $\hat{\sigma}_{n,1}(x) = \{\hat{\sigma}_1^2(x)\}^{1/2} n^{-1/2} h^{-1/2}$, where $\hat{\sigma}_1^2(x)$ is defined in (15). It is a consistent estimator of $\sigma_{n,1}^2(x)$ as $n \to \infty$ according to (10) and (16).

Similarly, according to (16), function $\sigma_{n,2}(x)$ with explicit expression in (37) is estimated consistently by

$$\hat{\sigma}_{n,2}(x) = \left\{\mathbf{\Delta}^T(x)\,\mathbf{S}_{j(x)}\mathbf{\Delta}(x)\right\}^{1/2}\left\{\hat{\sigma}_2^2(x)\right\}^{1/2}\{2nh/3\}^{-1/2},$$

with location index $j(x)$ defined in (5), 2×2 matrix $\mathbf{S}_j$ in (8), and $\hat{\sigma}_2^2(x)$ defined above, $\mathbf{\Delta}(x)$ defined as follows:

$$\mathbf{\Delta}(x) = \begin{pmatrix} c_{j(x)-1}\{1-\delta(x)\} \\ c_{j(x)}\delta(x) \end{pmatrix}, \quad c_j = \begin{cases} \sqrt{2} & j=-1, N \\ 1 & 0 \le j \le N-1 \end{cases}, \tag{18}$$

To express the pointwise variance of linear spline bands explicitly, one needs to find $\mathbf{S}_{j(x)}$, a 2×2 block matrix of the inverse of tridiagonal matrix $\mathbf{V}$. The solution for the inverse of a tridiagonal matrix can be derived via two theorems, Equation (43) in Gantmacher and Krein (1960) and Theorem 4.5 in Zhang (1999), please see Subsection 4.2, Wang and Yang (2009a) for details.

4 Simulation study

Simulated time series data is drawn from model (1) with

$$m(x) = \sin(2\pi x), \quad \sigma(x) = \sigma_0 \frac{100 - \exp(x-0.5)}{100 + \exp(x-0.5)},$$

with noise level $\sigma_0 = 0.5, 1.0$, sample size $n = 100, 200, 500$ and confidence level $1-\alpha = 0.95, 0.99$.

In this study the errors $\{\varepsilon_{in}\}_{i=1}^n$ are generated from an ARCH(1) model $\{v_i, i \ge 1\}$,

$$\begin{cases} v_i^2 = \omega_i^2 z_i^2, \; z_i \sim N(0,1) \\ \omega_i^2 = \alpha_0 + \alpha_1 v_{i-1}^2 \end{cases},$$

which is a popular member in martingale difference family. As one referee and the associate editor pointed out that a general ARCH model could not satisfy the error assumption (A3) directly. Note that ARCH(1) sequence have a constant unconditional variance $\alpha_0/(1-\alpha_1)$ given $\alpha_1 < 1$. Rescale the ARCH(1) sequence $\{v_i\}$ by its standard deviation is to obtain a martingale difference sequence with conditional mean 0 and unconditional variance 1. In particular, the sequence $\{v_i^2, i \geq 1\}$ follows AR(1) if $3\alpha_1^2 < 1$. Please see Fan and Yao (2003) and Schumway and Stoffer (2006) for details.

The error sequence of martingale differences $\{\varepsilon_{in}\}_{i=1}^n$ are generated as follows:

$$\varepsilon_{in} = \frac{v_i}{\sqrt{\alpha_0/(1-\alpha_1)}}, \quad 1 \leq i \leq n.$$

Simulation samples are firstly drawn from a ARCH model with coefficient α_0 fixed at 0.5. The degree of dependence depends on coefficient α_1 which is taken $\alpha_1 = 0, 0.1, 0.3, 0.5$. Typically $\alpha_1 = 0$ is associated with a serially uncorrelated sequence, and for normal errors in this study, it is an i.i.d. sequence. The larger the coefficient α_1, the stronger the dependence among the series $\{\varepsilon_{in}\}$.

As suggested in Song and Yang (2009), the bootstrap methods can improve the power and hence mean of bootstrap variance estimators is employed to model the pointwise variance. There are 200 Bootstrap samples randomly selected for all simulations. The Bayesian information criterion (BIC) is selected to optimize the knots selection in variance estimation. Because of the special piece-wise polynomial structure of splines, in particular jump type in constant spline, relatively more knots are selected in the constant spline estimation for variance estimation. While for linear spline type, BIC leads to a satisfactory coverage outcome with very limited knots number.

The empirical coverage probability are reported in Tables 1 and 2. If the simultaneous confidence bands constructed on the sample data as in (17) cover the true curve at every point in the range of [0, 1], the coverage counts 1; otherwise 0. In each simulation the number is counted out of 500 replications. The empirical coverage probability is calculated as the percentage out of 500 replications for comparison. All values outside (inside) the parenthesis are corresponding to coverage probability for spline bands with nominal confidence level 0.99(0.95).

Table 1 Constant spline bands coverage probabilities in 500 replications, with confidence 0.95 and 0.99 inside/outside of the parenthesis

σ_0	n	$\alpha_1 = 0$	$\alpha_1 = 0.1$	$\alpha_1 = 0.3$	$\alpha_1 = 0.5$
0.5	100	0.890 (0.630)	0.848 (0.576)	0.806 (0.570)	0.718 (0.498)
	200	0.924 (0.722)	0.906 (0.686)	0.868 (0.650)	0.824 (0.548)
	500	0.956 (0.752)	0.922 (0.726)	0.874 (0.710)	0.840 (0.574)
1.0	100	0.926 (0.746)	0.900 (0.668)	0.860 (0.670)	0.778 (0.588)
	200	0.968 (0.824)	0.934 (0.806)	0.904 (0.746)	0.864 (0.672)
	500	0.986 (0.898)	0.978 (0.884)	0.946 (0.838)	0.940 (0.762)

Table 2 Linear spline bands coverage probabilities in 500 replications, with confidence 0.95 and 0.99 inside/outside of the parenthesis

σ_0	n	$\alpha_1 = 0$	$\alpha_1 = 0.1$	$\alpha_1 = 0.3$	$\alpha_1 = 0.5$
0.5	100	0.996 (0.980)	0.996 (0.968)	0.948 (0.918)	0.858 (0.794)
	200	1.000 (0.978)	0.994 (0.978)	0.974 (0.944)	0.900 (0.852)
	500	1.000 (0.986)	1.000 (0.980)	0.992 (0.976)	0.954 (0.926)
1.0	100	0.988 (0.988)	0.988 (0.982)	0.938 (0.908)	0.894 (0.860)
	200	1.000 (1.000)	0.998 (0.992)	0.970 (0.958)	0.904 (0.888)
	500	1.000 (1.000)	1.000 (1.000)	1.000 (0.992)	0.958 (0.950)

Regardless of the noise level and dependence level, the confidence bands with confidence 0.99 always have better coverage than those with confidence 0.95, especially all the coverage probability are 100% for linear spline bands with confidence 0.99. With sample size increasing, there is noticeable improvement in corresponding coverage probability.

In both tables it is also found that confidence band with larger noise level 1.0 provide better coverage than those with noise level 0.5, which is quite intuitive since larger noise makes wider bands. Compared with i.i.d. cases ($\alpha_1 = 0$), confidence bands with weakly dependent errors have relatively smaller coverage probability.

In all cases, linear confidence bands outperform constant bands. The trigonometric function is differentiable to infinite order, so higher order spline produces high order smooth curves to fit the data better and capture the shape better. Compared with constant bands, linear bands could use fewer knots and have much better coverage. With fewer knots, it could significantly reduce the computation burden.

The linear spline band is recommended because of its stable statistical property and nice practical performance. In the following real data example at Sect. 5, I use the linear spline estimation and its confidence band to model time series trend.

To illustrate the proposed method, a sample with size 200 or 500 is drawn to create the graphs in Figs. 1 and 2. There are four types of symbols: points (data), solid smooth curve (true curve), thin dashed line (spline estimator), and upper and lower solid line (confidence bands). Compared with jump-type constant spline regression lines and bands on the left, those on the right corresponding to linear splines show better smoothness and capture the curve shape quite well. In all graphs, the confidence bands with larger confidence 0.99 is wider than those with smaller confidence 0.95. In addition, a larger sample produces narrower confidence bands.

5 Example of leaf area index

Many studies demonstrate the influence of land use and land cover change on local and regional climate. The climate and land use interaction project, or CLIP (http://clip.msu.edu) attempts to understand the nature and magnitude of the interactions of climate and land use/cover change across East Africa. For details of the project, please see Olson et al. (2008) and Wang et al. (2006). Phenological information reflecting

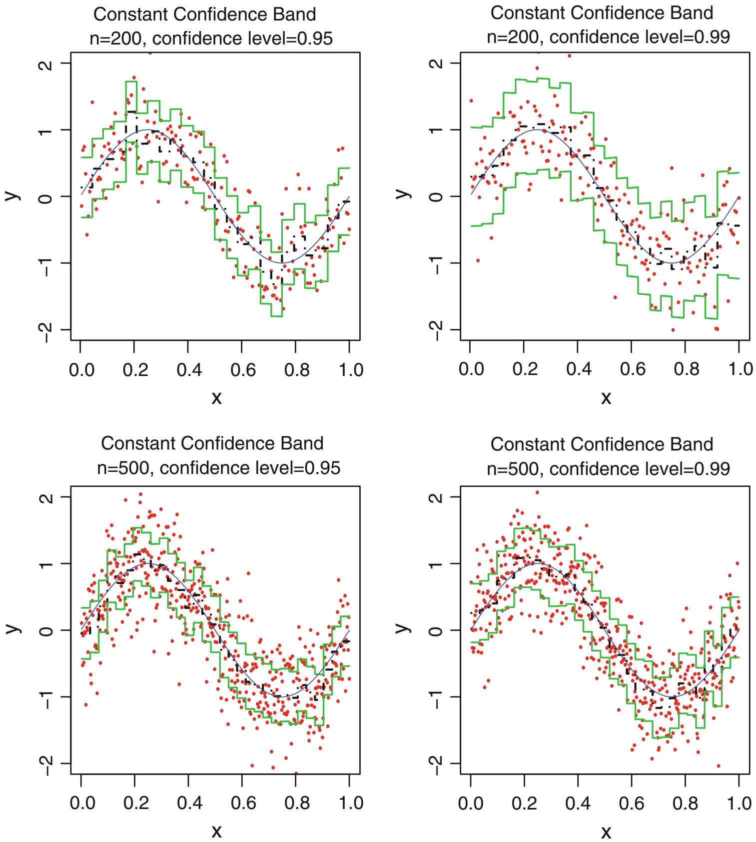

Fig. 1 Plots of constant spline bands, linear confidence bands (*solid lines*), linear spline estimators (*dot-dashed line*), simulated data (*dots*), and true curves (*thin line*)

the seasonal variability of vegetation is an important input variable in regional climate models such as the regional atmospheric modelling system (RAMS). It varies among different land cover types and geographic locations (latitude and longitude).

The RAMS version 4.4 is a state-of-the-art three-dimensional atmospheric model. For a given land cover class, it provides functions for several vegetation characteristics including LAI, fractional cover, roughness length, and displacement height. Although these characteristics are interrelated, in this paper only LAI is considered temporally and spatially for each land type.

The MODIS (Moderate Resolution Imaging Spectroradiometer) LAI product used is available at 8-day temporal intervals with a 1 km spatial resolution covering the entire study region in a two-dimensional tessellation. The data was obtained through the NASA Land Processes Distribution Active Archive Center. The phenological discrepancy between the RAMS model and the remote sensing measurement will show that the pre-assumed function relationship (in RAMS) is significantly different from the collected information of MODIS.

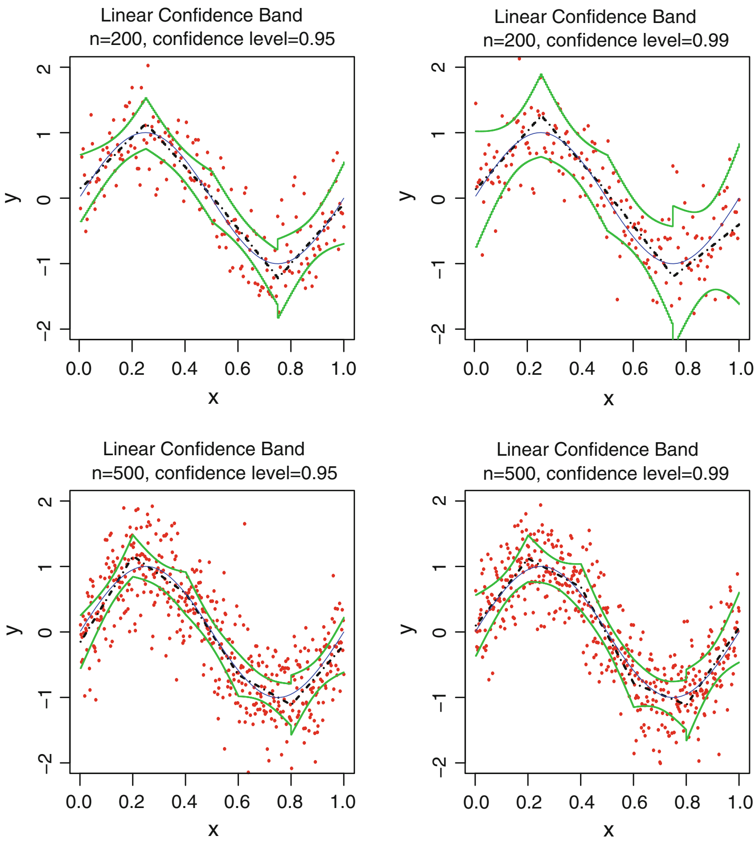

Fig. 2 Plots of linear spline bands, linear confidence bands (*solid lines*), linear spline estimators (*dot-dashed line*), simulated data (*dots*), and true curves (*thin line*)

To show the difference driven by the spatial affect, in particular the latitude, based on MODIS data, I have employed the linear spline method to estimate LAI function by formula (13) and its confidence bands in Theorem 1. The RAMS presumed function curve is only considered at three latitudes, North $5°(+5)$, and South $5°(-5)$. Each grid block covers the area of 0.1 by 0.1 degrees, the longitudinal of three grid blocks are chosen to be as close as possible. There are dozens of land cover types in East Africa region. For illustration purpose, two popular land cover types are selected for the study: deciduous woodland and deciduous shrubland with sparse trees.

To investigate the discrepancy, statistical hypotheses for each land type are set first: H_0 : LAI trend curve follows RAMS model vs. H_1: Does not follow RAMS model.

In Fig. 3, columns correspond to different latitudes, and rows for land cover types. The solid line represents the LAI hypothesized curve in the RAMS model, the central dashed line is the linear spline regression line, and the dotted lines (upper and lower) are the confidence bands derived from the MODIS data based on the proposed linear splines.

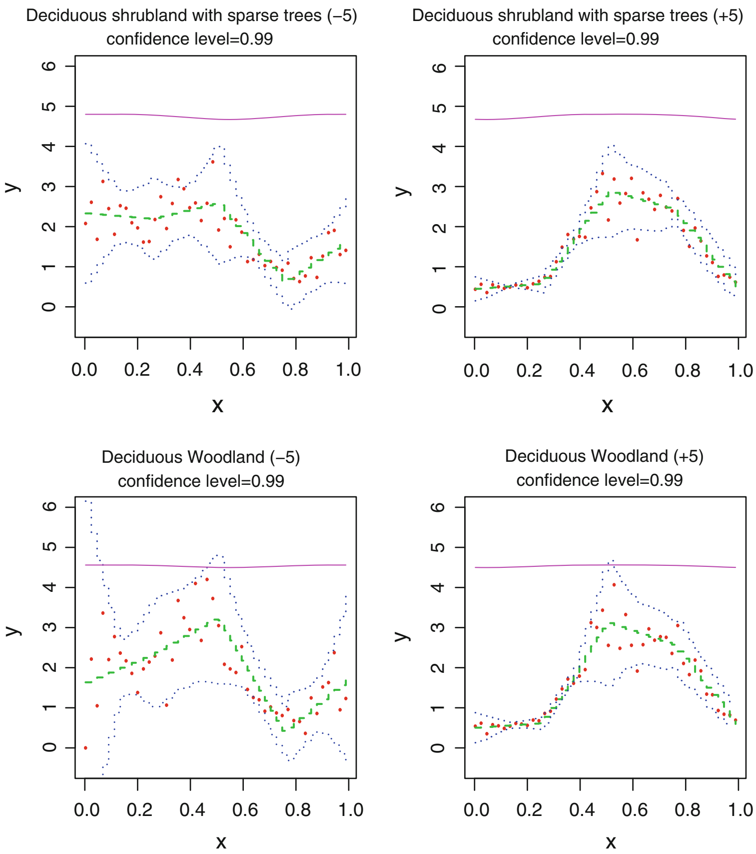

Fig. 3 Testing the time trend of leaf area index, linear confidence bands (*upper and lower dotted lines*), linear spline estimators (*central dashed line*), real data (*dots*), and hypothesized curves (*solid line*)

Given that the significance level as low as 0.01, the RAMS curve falls outside of the bands almost completely at every point at North 5 and South 5 latitude degrees. Therefore the test confirms that the RAMS curves overestimate the LAI, with the difference being significantly large indicated from the small p-value < 0.01. Applying the same methods to the data of other land cover types at different spatial locations yields similar results.

Appendix: Technical results

In the proofs, denote by $U(\cdot)$, $u(\cdot)$ quantities of orders $O(\cdot)$ and $o(\cdot)$ uniformly over $x \in [0, 1]$ and/or $1 - p \leq j \leq N + 1$.

Major technical results used in this section are the Bernstein's Inequality (Theorem 1.4 from Bosq (1996)), the Berry–Esseen Bound (Theorem 1 from Sunklodas (1984)), and the upcrossing probability result (Theorem 1.5.3 from

Leadbetter et al. (1983)). We denote by Φ the standard normal distribution function. For detailed proof of lemmas and theorems, please see Wang (2009).

Theorem 2 (Sunklodas 1984) *Let $\{\xi_i\}_{i=1}^n$ be an α-mixing sequence with $E\xi_i = 0$. Denote $d := \max_{1\le i\le n}\{E|\xi_i|^{2+\delta}\}$, $0 < \delta \le 1$, $S_n = \sum_{i=1}^n \xi_i$, $\sigma_n^2 := ES_n^2 \ge c_0 n$ for some $c_0 \in (0, +\infty)$. If $\alpha(n) \le K_0 e^{-\lambda_0 n}$, $\lambda_0 > 0$, $K_0 > 0$, then there exist $c_1 = c_1(K, \delta)$, $c_2 = c_2(K, \delta)$, such that*

$$\Delta_n = \sup_z \left| P\left\{\sigma_n^{-1} S_n < z\right\} - \Phi(z)\right| \le c_1 \frac{d}{c_0 \sigma_n^\delta}\left\{\log\left(\sigma_n / c_0^{1/2}\right)/\lambda\right\}^{1+\delta},$$

for any λ with $\lambda_1 \le \lambda \le \lambda_2$, where

$$\lambda_1 = c_2\left\{\log\left(\sigma_n / c_0^{1/2}\right)\right\}^b / n, \quad b > 2(1+\delta)/\delta; \quad \lambda_2 = 4\delta^{-1}\log\left(\sigma_n / c_0^{1/2}\right).$$

Recall that $G^{(-1)}$ is the space of functions that are constant on each J_j, and $G^{(0)}$ the space of functions that are linear on each J_j and continuous on $[0, 1]$. One could easily calculate the theoretical norms of their B-spline basis.

Lemma 1

$$\|b_{j,1}\|_2^2 = \int I_j^2(x)\,\mathrm{d}x = h, \quad j = 0, 1, \ldots, N,$$
$$\|b_{j,2}\|_2^2 = \int K^2\left\{\left(x - t_{j+1}\right)h^{-1}\right\}\mathrm{d}x = \begin{cases} 2h/3, & 0 \le j \le N-1 \\ h/3, & j = -1, N \end{cases},$$
$$\left\langle b_{j,2}, b_{j',2}\right\rangle = \int K\left(\frac{x - t_{j+1}}{h}\right)K\left(\frac{x - t_{j'+1}}{h}\right)\mathrm{d}x = \begin{cases} h/6, & \left|j - j'\right| = 1 \\ 0, & \left|j - j'\right| > 1 \end{cases}.$$

The next lemma connects theoretical and empirical norms. With the help of this lemma, the stochastic form (empirical) of inner product matrix can approximate the deterministic inner product, which provides an explicit expression to implement the proposed estimation in Sects. 3 and 4.

Lemma 2 *Under Assumption* (A5), *for $p = 1, 2$, as $n \to \infty$*

$$A_{n,p} = \sup_{g_1, g_2 \in G^{(p-2)}} \left| \frac{\langle g_1, g_2\rangle_n - \langle g_1, g_2\rangle}{\|g_1\|_2 \, \|g_2\|_2} \right| = O(n^{-1}h^{-1}) = o(1). \tag{19}$$

Proof It can be shown that when $p = 1$, $\max_{0\le j\le N} \|B_{j,1}\|_{2,n}^2 = 1 + O(n^{-1}h^{-1})$.

For any $g_1(x) \equiv \sum_{j=0}^{N} \lambda_{j1} B_{j,1}(x)$, $g_2(x) \equiv \sum_{j=0}^{N} \lambda_{j2} B_{j,1}(x)$,

$$\begin{aligned}
\left| \frac{\langle g_1, g_2 \rangle_n - \langle g_1, g_2 \rangle}{\|g_1\|_2 \|g_2\|_2} \right| &= \left| \frac{\sum_{j=0}^{N} \lambda_{j1} \lambda_{j2} \left(\| B_{j,1} \|_{2,n}^2 - 1 \right)}{\left\{ \sum_{j=0}^{N} \lambda_{j1}^2 \sum_{j=0}^{N} \lambda_{j2}^2 \right\}^{1/2}} \right| \\
&\leq \frac{\max_{0 \leq j \leq N} \left| \| B_{j,1} \|_{2,n}^2 - 1 \right| \left| \sum_{j=0}^{N} \lambda_{j1} \lambda_{j2} \right|}{\left\{ \sum_{j=0}^{N} \lambda_{j1}^2 \sum_{j=0}^{N} \lambda_{j2}^2 \right\}^{1/2}} \\
&\leq \max_{0 \leq j \leq N} \left| \| B_{j,1} \|_{2,n}^2 - 1 \right| = O\left(n^{-1} h^{-1} \right).
\end{aligned}$$

Proof for case $p = 2$ is similar to above procedures. □

A.1 Proof of Theorem 1 when $p = 1$

The projections of true function $\mathbf{m}(x)$ and pseudo function based on error $\mathbf{E}(i/n) \equiv \{\sigma(i/n)\varepsilon_{in}\}_{i=1}^{n}$ can be formulated as follows

$$\tilde{m}_p(x) = \left\{ B_{j,p}(x) \right\}_{1-p \leq j \leq N}^{T} \left(\left\langle B_{j',p}, B_{j,p} \right\rangle_n \right)_{1-p \leq j, j' \leq N}^{-1} \left\{ \left\langle \mathbf{m}, B_{j,p} \right\rangle_n \right\}_{j=1-p}^{n}, \tag{20}$$

$$\tilde{\varepsilon}_p(x) = \left\{ B_{j,p}(x) \right\}_{1-p \leq j \leq N}^{T} \left(\left\langle B_{j',p}, B_{j,p} \right\rangle_n \right)_{1-p \leq j, j' \leq N}^{-1} \left\{ \left\langle \mathbf{E}, B_{j,p} \right\rangle_n \right\}_{j=1-p}^{n}. \tag{21}$$

Note that $\tilde{\varepsilon}_1(x)$ is the error projected into the constant spline space with respect to empirical inner product, can be represented by the following

$$\tilde{\varepsilon}_1(x) = \sum_{j=0}^{N} \left\langle \mathbf{E}, B_{j,1} \right\rangle_n \| B_{j,1} \|_{2,n}^{-1} B_{j,1}(x). \tag{22}$$

Define $\varepsilon_j^* \equiv \langle \mathbf{E}, B_{j,1} \rangle_n = \frac{1}{n} \sum_{i=1}^{n} B_{j,1}(\frac{i}{n}) \sigma(\frac{i}{n}) \varepsilon_{in}$, a similar estimator $\hat{\varepsilon}_1(x)$ with $\|B_{j,1}\|_{2,n}$ replaced by $\|B_{j,1}\|_2$ in (22) can be simplified as

$$\hat{\varepsilon}_1(x) = \sum_{j=0}^{N} \varepsilon_j^* B_{j,1}(x) \| B_{j,1} \|_2^{-1} = \sum_{j=0}^{N} \varepsilon_j^* B_{j,1}(x), \quad x \in [0, 1]. \tag{23}$$

Later I will show that under Lemma 2, the difference between $\tilde{\varepsilon}_1(x)$ in (22) and $\hat{\varepsilon}_1(x)$ in (23) is asymptotically negligible and these two processes share a common asymptotic distribution.

Let us first work on the pointwise variance of $\hat{\varepsilon}_1(x)$.

Lemma 3 *Under Assumptions* (A1) *to* (A3) *and* (A5), *the pointwise variance of* $\hat{\varepsilon}_1(x)$ *can be approximated by* $\sigma^2(x)(nh)^{-1}$, *i.e.,*

$$E\left\{\hat{\varepsilon}_1(x)\right\}^2 = \sigma^2(x)(nh)^{-1}\left\{1 + r_{n,1}(x)\right\}, \quad x \in [0,1], \tag{24}$$

with $\sup_{x\in[0,1]} |r_{n,1}(x)| \to 0$ *as* $n \to \infty$. *In addition* $|E\{\hat{\varepsilon}_1(x)\}^2 - \sigma^2_{n,1}(x)| = u((nh)^{-1})$, *with* $\sigma^2_{n,1}(x)$ *defined in* (10).

Proof The main tools for the proof are Lemma 1 and continuity of function $\sigma(x)$. Note that $E\hat{\varepsilon}_1(x) = 0$. According to (23),

$$E\left\{\hat{\varepsilon}_1(x)\right\}^2 = \frac{I_{j(x)}(x)}{n^2h}\left\{\sum_{i=1}^{n} B^2_{j(x),1}\left(\frac{i}{n}\right)\sigma^2\left(\frac{i}{n}\right)\right\} = \sum_{\frac{i}{n}=hj(x)}^{h(j(x)+1)} \frac{\sigma^2\left(\frac{i}{n}\right)}{n^2h^2},$$

where the martingale difference sequence $\{\varepsilon_{in}, 1 \le i \le n\}$ is an uncorrelated sequence. Hence $\sup_{x\in[0,1]} |E\{\hat{\varepsilon}_1(x)\}^2 - \sigma^2(x)(nh)^{-1}|$

$$\le \sup_{x\in[0,1]} \frac{1}{n^2h^2}\left|\sum_{\frac{i}{n}=hj(x)}^{h(j(x)+1)} \sigma^2(i/n) - (nh)\sigma^2(x)\right| = o\left(n^{-1}h^{-1}\right),$$

which proves (24). Based on the continuity of the variance function, we have

$$\begin{aligned}&\sup_{x\in[0,1]}\left|B^2_{j(x),1}(i/n)\,\sigma^2(i/n) - B^2_{j(x),1}(x)\,\sigma^2(x)\right| \\ &\quad\le \sup_{x\in[0,1]} B^2_{j(x),1}(i/n)\,\omega(\sigma^2, h) = o(h^{-1})\end{aligned}$$

In addition, $\sup_{x\in[0,1]} |E\{\hat{\varepsilon}_1(x)\}^2 - \sigma^2_{n,1}(x)|$ is equal to

$$\sup_{x\in[0,1]} (nh)^{-1}\left|\sum_{\frac{i}{n}=hj(x)}^{h(j(x)+1)} \frac{\sigma^2(i/n)}{nh} - \int I_{j(x)}(v)\,\sigma^2(v)\,\mathrm{d}v\right| = o\left((nh)^{-1}\right).$$

□

Next lemma proves that the difference between $\tilde{\varepsilon}_1(x)$ in (22) and $\hat{\varepsilon}_1(x)$ in (23) is negligible uniformly over $x \in [0,1]$, which is directly derived from (19).

Lemma 4 *Under Assumptions* (A1) *to* (A3) *and* (A5),

$$\left|\tilde{\varepsilon}_1(x) - \hat{\varepsilon}_1(x)\right| \le A_{n,1}\left(1 - A_{n,1}\right)^{-1}\left|\hat{\varepsilon}_1(x)\right|, \quad x \in [0,1].$$

The Berry–Esseen Theorem in Sunklodas (1984) is a critical tool to the main technical results in this paper. Major assumptions in this paper are on those moments conditions: the second moment of a α-mixing sequence sum with order n, and finite $(2+\delta)$ moment. For simplicity, take $\delta = 1$. To apply the Berry–Esseen result, one needs to verify the moment conditions for a new defined sequence

$$\xi_{i,j} \equiv B_{j,1}(i/n)\,\sigma(i/n)\,\varepsilon_i, \quad 1 \le i \le n, \quad j = 0, \ldots, N.$$

Lemma 5 *Under Assumptions* (A1) *to* (A3)*, there exist constants* $c_{\sigma,0}, C_{\sigma,0} \in (0, +\infty)$*, such that for each* $j = 0, \ldots, N$

$$\sigma_j^2 = E\left(\sum_{i=1}^n \xi_{i,j}\right)^2 = nc_{\sigma,j} \ge nc_{\sigma,0}, \tag{25}$$

$$d_j = E|\xi_{i,j}|^3 = E\left\{B_{j,1}^3(i/n)\,\sigma^3(i/n)\,|\varepsilon_{in}|^3\right\} \le C_{\sigma,0}h^{-3/2}. \tag{26}$$

Proof For martingale difference sequence $\{\varepsilon_{in},\}_{i=1}^n$ we have $E\{\varepsilon_{in}\varepsilon_{i'n}\}=0$ for $i \ne i'$, and $E\{\varepsilon_{in}^2\} = 1$, then $\forall 0 \le j \le N$,

$$c_{\sigma,j} = \sigma_j^2/n = (nh)^{-1} \sum_{jh \le \frac{i}{n} \le (j+1)h} \sigma^2\left(\frac{i}{n}\right) \ge c_{\sigma,0} > 0.$$

Based on Lemma 1 and the continuity of functions $\sigma^2(x)$, it implies that

$$E|\xi_{i,j}|^3 \le B_{j,1}^3(i/n)\,\sigma^3(i/n)\,E|\varepsilon_i|^3 \le h^{-3/2}C_{\sigma,0},$$

where $C_{\sigma,0} = \sup_{0 \le x \le 1}\{\sigma^3(x)E|\varepsilon_i|^3\} < \infty$ since $\max_{1 \le i \le n}\{E|\varepsilon_i|^3\} < \infty$ given in Assumption (A3). □

Lemma 6

$$\liminf_{n\to\infty} P\left[\sup_{x\in[0,1]}\left|\sigma_{n,1}^{-1}(x)\,\tilde{\varepsilon}_1(x)\right| \le \{2\log(N+1)\}\,d_n(\alpha)\right] \ge 1 - \alpha$$

Proof For any $x \in [0,1]$, the standardized term $\hat{\varepsilon}_1(x) = h^{-1/2}n^{-1}\sum_{i=1}^n \xi_{i,j(x)}$ can be reexpressed as $(E\{\hat{\varepsilon}_1^2(x)\})^{-1/2}\hat{\varepsilon}_1(x) = \sigma_{j(x)}^{-1}\sum_{i=1}^n \xi_{i,j(x)}$. Then one has that $|\sigma_{n,1}^{-1}(x)\hat{\varepsilon}_1(x) - \sigma_{j(x)}^{-1}\sum_{i=1}^n \xi_{i,j}|$ equals to

$$\left|\left[\sigma_{n,1}^{-1}(x) - \left(E\left\{\hat{\varepsilon}_1^2(x)\right\}\right)^{-1/2}\right] B_{j(x),1}(x)\,\frac{1}{n}\sum_{i=1}^n \xi_{i,j(x)}\right| = u((nh)^{-3/2}) = u(1)$$

since $E\{\frac{1}{n}\sum_{i=1}^n \xi_{i,j}\}^2 = n^{-2}\sigma_j^2 \le n^{-1}C_\sigma^2$ implies that $\frac{1}{n}\sum_{i=1}^n \xi_{i,j} = U(n^{-1/2})$, in which $\sigma_j^2 = nc_{\sigma,j}$, $C_\sigma = \max_{0\le x\le 1}\{\sigma(x)\}$, and $d_j \le C_{\sigma,0}h^{-3/2}$ as in (26).

The sequence $\{\xi_{i,j}\}_{i=1}^{n}$ is a stationary α-mixing sequence, with mean $E\xi_{i,j} = 0$. Thus by Theorem 2 with δ replaced by 1, there exist positive constants c_1 and c_2 such that

$$\begin{aligned}\Delta_n &= \max_{0\le j\le N}\sup_z\left|P\left\{\frac{\frac{1}{n}\sum_{i=1}^{n} B_{j,1}(x)\,B_{j,1}(i/n)\,\sigma(i/n)\,\varepsilon_{in}}{\left[E\left\{\hat{\varepsilon}_1(x)\right\}^2\right]^{1/2}}\le z, x\in I_j\right\}-\Phi(z)\right|\\ &= \max_{0\le j\le N}\sup_z\left|P\left\{\sigma_{j(x)}^{-1}\sum_{i=1}^{n}\xi_{i,j(x)}\le z\right\}-\Phi(z)\right|+o(1)\\ &\le c_1\frac{\left\{C_{\sigma,0}h^{-3/2}\right\}}{c_{\sigma,0}\sigma_{j(x)}}\left\{\log\left(\sigma_{j(x)}/\sqrt{c_{\sigma,0}}\right)/\lambda\right\}^2.\end{aligned}$$

For any $c_2.n^{-1}\{\log(\sigma_{j(x)}/\sqrt{c_{\sigma,0}})\}^b \le \lambda \le 12\{\log(\sigma_{j(x)}/\sqrt{c_{\sigma,0}})\}$, $b > 4$, then

$$\Delta_n \le c_1\frac{\left\{C_{\sigma,0}h^{-3/2}\right\}}{c_{\sigma,0}\left\{nc_{\sigma,0}\right\}^{1/2}}12^{-2} \le C\left(nh^3\right)^{-1/2}.$$

Assumption (A5) leads to $\Delta_n \to 0$ uniformly as $n \to \infty$.

Define

$$a_N=(2\log N)^{1/2},\, b_N=(2\log N)^{1/2}-\frac{(2\log N)^{-1/2}(\log\log N+\log 4\pi)}{2}. \tag{27}$$

Based on Theorem 1.5.3 in Leadbetter et al. (1983), one has

$$\begin{aligned}&P\left\{\sup_{0\le j\le N}\left|\sum_{i=1}^{n}\frac{\xi_{i,j}}{\sigma_j}\right|>\frac{-\log(\alpha/2)}{a_{N+1}}+b_{N+1}\right\}\\ &\quad\le\sum_{j=0}^{N}P\left\{\left|\sum_{i=1}^{n}\frac{\xi_{i,j}}{\sigma_j}\right|>\frac{-\log(\alpha/2)}{a_{N+1}}+b_{N+1}\right\}=\alpha+U\left(\left(nh^3\right)^{-1/2}\right)\end{aligned}$$

Let $-\log(\alpha/2)/a_{N+1} + b_{N+1} = \{2\log(N+1)\}^{1/2}d_n(\alpha)$, i.e.

$$d_n(\alpha) = 1-\{2\log(N+1)\}^{-1/2}\left[\log(\alpha/2)+\frac{\log\log(N+1)+\log 4\pi}{2}\right]. \tag{28}$$

Therefore,

$$\begin{aligned}
&\liminf_{n\to\infty} P\left[\sup_{x\in[0,1]}\left|\sigma_{n,1}^{-1}(x)\,\hat{\varepsilon}_1(x)\right| \le \{2\log(N+1)\}^{1/2} d_n(\alpha)\right] \\
&\quad = \liminf_{n\to\infty} P\left[\sup_{0\le j\le N}\left|\sigma_j^{-1}\sum_{i=1}^{n}\xi_{i,j}\right| \le \{2\log(N+1)\}^{1/2} d_n(\alpha)\right] \ge 1-\alpha.
\end{aligned}$$

□

Proof of Proposition 1 Based on Lemma 4, we have

$$\left|\sup_{x\in[0,1]}\left|\sigma_{n,1}^{-1}(x)\hat{\varepsilon}_1(x)\right| - \sup_{x\in[0,1]}\left|\sigma_{n,1}^{-1}(x)\,\tilde{\varepsilon}_1(x)\right|\right| = O_p\left((nh)^{-1}\right) = o_p(1).$$

It implies that $\sup_{x\in[0,1]}|\sigma_{n,1}^{-1}(x)\hat{\varepsilon}_1(x)|$ share the same asymptotic distribution as $\sup_{x\in[0,1]}|\sigma_{n,1}^{-1}(x)\tilde{\varepsilon}_1(x)|$. Hence

$$\liminf_{n\to\infty} P\left[\sup_{x\in[0,1]}\left|\sigma_{n,1}^{-1}(x)\,\tilde{\varepsilon}_1(x)\right| \le \{2\log(N+1)\}^{1/2} d_n(\alpha)\right] \ge 1-\alpha.$$

□

Proof of Theorem 1 From the approximation theorem in Pg. 154 in de Boor (2001) and Theorem 5.1 in Huang (2003), we have $\|\tilde{m}_1(x) - m(x)\|_\infty = O_p(h)$. Since $h \ll (nh)^{-1/2}\log^{1/2}(N+1)$ according to Assumption (A5), the bias order is negligible compared to the noise order. Applying Proposition 1,

$$\begin{aligned}
&\liminf_{n\to\infty} P\left[m(x) \in \hat{m}_1(x) \pm \sigma_{n,1}(x)\{2\log(N+1)\}^{1/2} d_n(\alpha),\, \forall x \in [0,1]\right] \\
&= \liminf_{n\to\infty} P\left[\sup_{x\in[0,1]} \sigma_{n,1}^{-1}(x)\left|\tilde{\varepsilon}_1(x)+\tilde{m}_1(x)-m(x)\right| \le \{2\log(N+1)\}^{1/2} d_n(\alpha)\right] \\
&= \liminf_{n\to\infty} P\left[\sup_{x\in[0,1]}\left|\sigma_{n,1}^{-1}(x)\,\tilde{\varepsilon}_1(x)\right| \le \{2\log(N+1)\}^{1/2} d_n(\alpha)\right] \ge 1-\alpha.
\end{aligned}$$

□

A.2 Proof of Theorem 1 when $p = 2$

In this subsection we examine some matrices used to construct linear confidence band in Theorem 1. In what follows, we use $|T|$ to denote the maximal absolute value of any matrix T.

Lemma 7 *The inner product matrix* $\mathbf{V}$ *of the B-spline basis* $\{B_{j,2}(x)\}_{j=-1}^{N}$ *defined in* (7) *can be calculated*

$$\mathbf{V} = \left(\langle B_{j',2}, B_{j,2}\rangle\right)_{j,j'=-1}^{N} = \begin{pmatrix} 1 & \sqrt{2}/4 & 0 & & & 0 \\ \sqrt{2}/4 & 1 & 1/4 & \ddots & & \\ 0 & 1/4 & 1 & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & 1/4 & 0 \\ & & \ddots & 1/4 & 1 & \sqrt{2}/4 \\ 0 & & & 0 & \sqrt{2}/4 & 1 \end{pmatrix}. \tag{29}$$

Consider matrix $\mathbf{S} = (s_{j'j})_{j',j=-1}^{N} \equiv \mathbf{V}^{-1}$, let $\tilde{\boldsymbol{\xi}}_{j'} = \{\mathrm{sgn}(s_{j'j})\}_{j=-1}^{N}$, apply Lemma B.2 in the Supplement of Wang and Yang (2009a), then there exists a positive C_s such that

$$\sum_{j=-1}^{N} \left|s_{j'j}\right| \le \left|\mathbf{S}\tilde{\boldsymbol{\xi}}_{j'}\right| \le C_s \left|\tilde{\boldsymbol{\xi}}_{j'}\right| = C_s, \quad \forall j' = -1, 0, \ldots, N. \tag{30}$$

The linear spline estimator of error term $\tilde{\varepsilon}_2(x)$ in (21) can be expressed as

$$\tilde{\varepsilon}_2(x) = \sum_{j=-1}^{N} \tilde{a}_j B_{j,2}(x), \quad x \in [0,1], \tag{31}$$

where the linear spline coefficient vector $\tilde{\mathbf{a}} = (\tilde{a}_{-1}, \ldots, \tilde{a}_N)^T$ is defined as

$$\tilde{\mathbf{a}} = \left(\mathbf{V} + \tilde{\mathbf{B}}\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} \xi_{i,j}^{L}\right)_{j=-1}^{N},$$

where $\xi_{i,j}^{L} \equiv B_{j,2}(\frac{i}{n})\sigma(\frac{i}{n})\varepsilon_i$ and term $\tilde{\mathbf{B}}$ satisfies $|\tilde{\mathbf{B}}| \le A_{n,2} = O(n^{-1}h^{-1})$ by Lemma 2.

Now define $\hat{\mathbf{a}} = (\hat{a}_{-1}, \ldots, \hat{a}_N)^T$ by replacing $(\mathbf{V} + \tilde{\mathbf{B}})^{-1}$ with $\mathbf{V}^{-1} = \mathbf{S}$ in above formula, i.e.

$$\hat{\mathbf{a}} = \mathbf{S}\left\{\frac{1}{n}\sum_{i=1}^{n} \xi_{i,j}^{L}\right\}_{j=-1}^{N} = \left\{\sum_{j=-1}^{N} s_{j'j}\frac{1}{n}\sum_{i=1}^{n} \xi_{i,j}^{L}\right\}_{j'=-1}^{N} = \frac{1}{\sqrt{n}}\mathbf{S}\xi,$$

with

$$\xi \equiv \left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \xi_{i,j}^{L}\right\}_{j=-1}^{N}. \tag{32}$$

Recall that the location index $j(x)$ defined in (5) and $B_{j,2}(x)$ in (6), $\hat{\varepsilon}_2(x)$ can be expressed as

$$\hat{\varepsilon}_2(x) = \sum_{j'=-1}^{N} \hat{a}_{j'} B_{j',2}(x) = \sum_{j'=j(x)-1}^{j(x)} \hat{a}_{j'} B_{j',2}(x)$$
$$= \sum_{j'=j(x)-1}^{j(x)} \frac{1}{\sqrt{n}} B_{j',2}(x) \sum_{j=-1}^{N} \left\{ s_{j',j} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{i,j}^{L} \right\}. \tag{33}$$

which is to approximate $\tilde{\varepsilon}_2(x)$ in (31). For simplicity define vectors $\mathbf{D}(x)$ and $\{\Lambda_j\}_{j=0}^{N}$

$$\mathbf{D}(x) \equiv \frac{1}{\sqrt{n}} \begin{pmatrix} B_{j(x)-1,2}(x) \\ B_{j(x),2}(x) \end{pmatrix}, \quad \Lambda_j = \begin{pmatrix} \Lambda_{j1} \\ \Lambda_{j2} \end{pmatrix} \equiv \mathcal{S}_j \xi, \tag{34}$$

in which $\mathcal{S}_j$ is a $2 \times (N+2)$ matrix with the $(j-1)$th and jth rows of matrix $\mathbf{S}$ only

$$\mathcal{S}_j = \begin{pmatrix} s_{j-1,-1} & s_{j-1,0} & \cdots & s_{j-1,N} \\ s_{j,-1} & s_{j,0} & \cdots & s_{j,N} \end{pmatrix}, \quad 0 \le j \le N. \tag{35}$$

Then, one can write $\hat{\varepsilon}_2(x)$ in (33) by the following matrix form

$$\hat{\varepsilon}_2(x) = \mathbf{D}^T(x) \Lambda_{j(x)}, \quad x \in [0, 1], \tag{36}$$

Lemma 8 *The pointwise variance of* $\hat{\varepsilon}_2(x)$ *can be approximated by the function* $\sigma_{n,2}^2(x)$ *defined in* (11)*, which satisfies*

$$\sigma_{n,2}^2(x) = \frac{3\sigma^2(x)}{2nh} \mathbf{\Delta}^T(x) \mathbf{S}_{j(x)} \Delta(x) \left\{1 + r_{n,2}(x)\right\}, \tag{37}$$

with $\sup_{x\in[0,1]} |r_{n,2}(x)| \to 0$, $j(x)$ *is as defined in* (5)*,* $\mathbf{\Delta}(x)$ *as defined in* (18) *and matrix* $\mathbf{S}_j$ *in* (8)*. And further*

$$\left| E\left\{\hat{\varepsilon}_2^2(x)\right\} - \sigma_{n,2}^2(x) \right| = u(n^{-1}h^{-1})$$

Consequently, there exist positive constants c_σ, C_σ *such that for* n *large enough*

$$c_\sigma (nh)^{-1/2} \le \sigma_{n,2}(x) \le C_\sigma (nh)^{-1/2}, \quad \forall x \in [0, 1]. \tag{38}$$

Proof From (34) and (36), $E\{\hat{\varepsilon}_2^2(x)\}$ is equal to

$$\mathbf{D}^T(x) \operatorname{cov}\left(\Lambda_{j(x)}\right) \mathbf{D}(x) = \mathbf{D}^T(x) \mathcal{S}_{j(x)} \operatorname{cov}(\xi) \mathcal{S}_{j(x)}^T \mathbf{D}(x).$$

Note that $\{\varepsilon_{in}\}$ is a martingale difference sequence with $E(\varepsilon_{in}) = 0$, $E[\varepsilon_{in}\varepsilon_{kn}] = 0, \forall i \neq k$, the jlth entry of the covariance matrix $\mathrm{cov}(\xi)$ is

$$u_{jl} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n} E\left\{\xi_{i,j}^{L}\xi_{l,k}^{L}\right\} = \frac{1}{n}\sum_{i=1}^{n} B_{j,2}\left(\frac{i}{n}\right) B_{l,2}\left(\frac{i}{n}\right)\sigma^{2}\left(\frac{i}{n}\right)$$
$$= \int \sigma^{2}(v)\, B_{j,2}(v)\, B_{l,2}(v)\, \mathrm{d}v + u(1) = \sigma_{jl}^{2} + u(1)$$

where σ_{jl}^2 is the jlth entry of the matrix Σ defined in (9), i.e., $|\,\mathrm{cov}(\xi) - \Sigma| = \max_{-1\leq k,l\leq N} |u_{kl} - \sigma_{kl}^2| = o(1)$.

Simple matrix computation leads to

$$\mathcal{S}_{j(x)}\mathbf{D}(x) = \left(n^{-1/2}\left\{B_{j(x)-1}(x)\, s_{j(x)-1,j} + B_{j(x)}(x)\, s_{j(x),j}\right\}\right)_{j=-1}^{N}$$

and $\mathbf{D}^{T}(x)\mathcal{S}_{j(x)}\Sigma\mathcal{S}_{j(x)}^{T}\mathbf{D}(x)$ is the matrix product format of $\sigma_{n,2}^2(x)$ in (11). Hence it implies that $|E\{\hat{\varepsilon}_2^2(x)\} - \sigma_{n,2}^2(x)| = u(n^{-1}h^{-1})$.

The rest of proof is similar to Lemma B.4 in Wang and Yang (2009a). $\square$

Lemma 9 *Under Assumptions* (A1) *to* (A5)*,*

$$\max_{0\leq j\leq N}\left|\mathrm{cov}(\Lambda_j) - \sigma^2\left(t_{j+1}\right)\mathbf{S}_j\right| = o(1).$$

Proof Given $\mathcal{S}_j$ in (35), the covariance matrix of Λ_j is expressed as

$$\mathrm{cov}(\Lambda_j) = \mathrm{cov}(\mathcal{S}_j\xi) = \mathrm{cov}\left\{\frac{1}{\sqrt{n}}\begin{pmatrix}\sum_{i=1}^{n}\sum_{j'=-1}^{N} s_{j-1,j'}\xi_{i,j'}^{L}\\ \sum_{i=1}^{n}\sum_{j'=-1}^{N} s_{j,j'}\xi_{i,j'}^{L}\end{pmatrix}\right\}$$

Following the positive definite property of matrix $\mathbf{S}$ in Lemma B.2 in Supplement of Wang and Yang (2009a),

$$R_{j,1} = \left|\mathrm{cov}(\Lambda_j) - \mathcal{S}_j\Sigma\mathcal{S}_j^{T}\right|_{\infty} \leq C\left|u_{kl} - \sigma_{kl}^2\right| = u(1).$$
$$\mathcal{S}_j\Sigma\mathcal{S}_j^{T} = \begin{pmatrix}\sum_{k,l=-1}^{N} s_{j-1,k}\sigma_{kl}^2 s_{j-1,l} & \sum_{k,l=-1}^{N} s_{j-1,k}\sigma_{kl}^2 s_{j,l}\\ \sum_{k,l=-1}^{N} s_{j,k}\sigma_{kl}^2 s_{j-1,l} & \sum_{k,l=-1}^{N} s_{j,k}\sigma_{kl}^2 s_{j,l}\end{pmatrix} \tag{39}$$

The variance element in above matrix

$$\sigma_{kl}^2 = \int \sigma^2(v)\, B_{k,2}(v)\, B_{l,2}(v)\, \mathrm{d}v = \sigma^2(t_{k+1})\, v_{kl} + o\left(w\left(\sigma^2, h\right)\right)$$
$$= \sigma^2(t_{l+1})\, v_{kl} + o\left(w\left(\sigma^2, h\right)\right).$$

in which v_{kl} is the element of the inner product matrix of the linear spline basis function, i.e. $\mathbf{V} = (v_{kl})_{-1\le k,l\le N}$. Replace σ_{kl}^2 with $\sigma^2(t_{l+1})v_{kl}$ in (39), and denote

$$\Gamma_{j,j'} = \sum_{k,l=-1}^{N} s_{j,k}\sigma^2\left(t_{l+1}\right)v_{kl}s_{j',l} \quad \text{for } j, j' = 1, \ldots, N.$$

Based on the continuity assumption of the variance function $\sigma^2(\cdot)$, it is quite straightforward that the difference is also negligible

$$R_{j,2} = \left|\mathcal{S}_j \Sigma \mathcal{S}_j^T - \begin{pmatrix} \Gamma_{j-1,j-1} & \Gamma_{j-1,j} \\ \Gamma_{j,j-1} & \Gamma_{j,j} \end{pmatrix}\right|_\infty = u\left(1\right)$$

Since $\mathbf{S} = \mathbf{V}^{-1} = (v_{kl})^{-1}$, we have $\sum_{k=-1}^{N} s_{j,k}v_{kl} = 1$ when $j = l$, otherwise 0. So $\Gamma_{j-1,j} = \sigma^2(t_j)s_{j-1,j}$, and $\Gamma_{j,j} = \sigma^2(t_{j+1})s_{j,j}$, hence

$$\begin{aligned} R_{j,3} &= \left|\begin{pmatrix} \Gamma_{j-1,j-1} & \Gamma_{j-1,j} \\ \Gamma_{j,j-1} & \Gamma_{j,j} \end{pmatrix} - \sigma^2\left(t_{j+1}\right)\begin{pmatrix} s_{j-1,j} & s_{j-1,j} \\ s_{j,j-1} & s_{j,j} \end{pmatrix}\right| \\ &\le C_s\left|\sigma^2\left(t_j\right) - \sigma^2\left(t_{j+1}\right)\right| = u\left(h\right) = u\left(1\right) \end{aligned}$$

It is clear that

$$\left|\mathrm{cov}(\Lambda_j) - \sigma^2\left(t_{j+1}\right)\mathbf{S}_j\right| \le R_{j,1} + R_{j,2} + R_{j,3} = u\left(1\right).$$

which implies that $\sigma^2(t_{j+1})\mathbf{S}_j$ approximations to covariance $\mathrm{cov}(\Lambda_j)$ uniformly. Hence, the lemma follows. □

Define $\{\mathrm{cov}(\Lambda_j)\}^{-1/2} = \begin{pmatrix} \lambda_{11,j} & \lambda_{12,j} \\ \lambda_{21,j} & \lambda_{22,j} \end{pmatrix}$ and $\Lambda_j \equiv \mathcal{S}_j\xi$ as defined in (34). For any $0 \le j \le N$,

$$\left\{\mathrm{cov}(\Lambda_j)\right\}^{-1/2}\Lambda_j = \begin{pmatrix} \lambda_{11,j}\Lambda_{j1} + \lambda_{12,j}\Lambda_{j2} \\ \lambda_{21,j}\Lambda_{j1} + \lambda_{22,j}\Lambda_{j2} \end{pmatrix} = \left\{\mathrm{cov}(\Lambda_j)\right\}^{-1/2}\mathcal{S}_j\xi,$$

$$\Lambda_j^T\left\{\mathrm{cov}\left(\Lambda_j\right)\right\}^{-1}\Lambda_j = \sum_{l=1,2}\left\{\lambda_{l1,j}\Lambda_{j1} + \lambda_{l2,j}\Lambda_{j2}\right\} = \sum_{l=1,2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_{i,j,l}$$

where $\eta_{i,j,l} = \sum_{j'=-1}^{N}(\lambda_{l1,j}s_{j-1,j'} + \lambda_{l2,j}s_{j,j'})\xi_{i,j'}^L$, $l = 1, 2$.

Lemma 10 *Under Assumptions* (A1) *to* (A4),

$$\lim\sup_{n\to\infty} P\left\{\max_{0\le j\le N}\left|n^{-1/2}\sum_{i=1}^{n}\eta_{i,j,l}\right|^2 > 2\log\left(N+1\right)\left\{d_n\left(\frac{\alpha}{2}\right)\right\}^2\right\} = \frac{\alpha}{2}.$$

Proof For simplicity we only prove the case of $l = 1$, and without specific notation, $\eta_{i,j} = \eta_{i,j,1}$. Case of $l = 2$ can be derived similarly.

It is clear that $E\eta_{i,j} = 0$, $E(\sum_{i=1}^{n} \eta_{i,j})^2 = nE(\lambda_{11,j}\Lambda_{j1} + \lambda_{12,j}\Lambda_{j2})^2 = n$. Based on equations (30) and (38), and Lemma 9, the third moment

$$E\left|\eta_{i,j}\right|^3 \leq Ch^{-3/2}\sup\left\{E\left|\varepsilon\right|^3\right\} \leq C_T M_0 h^{-3/2}.$$

Then by Sunklodas lemma,

$$\Delta_n = \max_{j=-1,\ldots,N}\sup_z\left|P\left\{n^{-1/2}\sum_{i=1}^{n}\eta_{i,j} < z\right\} - \Phi(z)\right| = O\left(\left(nh^3\right)^{-1/2}\log^2 n\right).$$

Following similar procedure to the constant case, one has

$$\begin{aligned} &P\left\{\max_{0\leq j\leq N}\left|n^{-1/2}\sum_{i=1}^{n}\eta_{i,j}\right| > \{2\log(N+1)\}^{1/2} d_n\left(\frac{\alpha}{2}\right)\right\} \\ &\leq \frac{\alpha}{2} + O\left(\left(nh^5\right)^{-1/2}\log^2 n\right) + u(1) \end{aligned}$$

where d_n is defined in (28). Hence

$$\limsup_{n\to\infty} P\left\{\max_{0\leq j\leq N}\left|n^{-1/2}\sum_{i=1}^{n}\eta_{i,j}\right|^2 > 2\log(N+1)\, d_n^2\left(\frac{\alpha}{2}\right)\right\} = \frac{\alpha}{2}.$$

□

Lemma 11 *Under Assumptions* (A1) *to* (A5)*,*

$$\liminf_{n\to\infty} P\left[\sup_{x\in[0,1]}\left|\sigma_{n,2}^{-1}(x)\hat{\varepsilon}_2(x)\right| \leq 2\{\log(N+1)\}^{1/2} d_n(\alpha/2)\right] \geq 1-\alpha.$$

Proof Note that $\hat{\varepsilon}_2(x) = \mathbf{D}^T(x)\Lambda_{j(x)}$, where $\mathbf{D}(x)$ and $\Lambda_{j(x)}$ are defined in (34). Thus, standardization leads to

$$\left\{\sigma_{n,2}^{-1}(x)\hat{\varepsilon}_2(x)\right\}^2 = \frac{\mathbf{D}(x)^T\Lambda_{j(x)}\Lambda_{j(x)}^T\mathbf{D}(x)}{\mathbf{D}(x)^T\operatorname{cov}\left(\Lambda_{j(x)}\right)\mathbf{D}(x)} \leq \Lambda_{j(x)}^T\left\{\operatorname{cov}\left(\Lambda_{j(x)}\right)\right\}^{-1}\Lambda_{j(x)},$$

where the inequality is a direct application of maximization lemma of Johnson and Wichern (1992, page 166). It is obvious that

$$\sup_{x\in[0,1]}\left|\sigma_{n,2}^{-1}(x)\hat{\varepsilon}_2(x)\right|^2 \leq \max_{0\leq j\leq N}\left\{\Lambda_j^T\left\{\operatorname{cov}\left(\Lambda_j\right)\right\}^{-1}\Lambda_j\right\}.$$

Thus

$$\begin{aligned}
&\liminf_{n\to\infty} P\left[\sup_{x\in[0,1]}\left|\sigma_{n,2}^{-1}(x)\hat{\varepsilon}_2(x)\right| \le 2\left\{\log\left(N+1\right)\right\}^{1/2} d_n\left(\alpha/2\right)\right] \\
&\ge \liminf_{n\to\infty} P\left[\max_{0\le j\le N}\left\{\Lambda_j^T\left\{\operatorname{cov}\left(\Lambda_j\right)\right\}^{-1}\Lambda_j\right\} \le 4\left\{\log\left(N+1\right)\right\}\left\{d_n\left(\alpha/2\right)\right\}^2\right] \\
&\ge 1-\limsup_{n\to\infty} P\left[\max_{0\le j\le N}\sum_{l=1,2}\left\{n^{-1/2}\sum_{i=1}^{n}\eta_{i,j,l}\right\}^2 > 4\left\{\log\left(N+1\right)\right\}\left\{d_n\left(\alpha/2\right)\right\}^2\right] \\
&\ge 1-\sum_{l=1,2}\limsup_{n\to\infty} P\left[\max_{0\le j\le N}\left\{n^{-1/2}\sum_{i=1}^{n}\eta_{i,j,l}\right\}^2 > 2\left\{\log\left(N+1\right)\right\}\left\{d_n\left(\alpha/2\right)\right\}^2\right] \\
&\ge 1-\alpha/2\times 2 = 1-\alpha.
\end{aligned}$$

The last inequality derived from Lemma 10. □

Lemma 12 *Under Assumptions* (A1) *to* (A5)*, we have*

$$\left|\sup_{x\in[0,1]}\left|\sigma_{n,2}^{-1}(x)\hat{\varepsilon}_2\left(x\right)\right| - \sup_{x\in[0,1]}\left|\sigma_{n,2}^{-1}\left(x\right)\tilde{\varepsilon}_2\left(x\right)\right|\right| = O_p\left(n^{-1}h^{-1}\right) = o_p\left(1\right).$$

Proof Similar proof as in Lemma B.11 in Wang and Yang (2009a). □

Proof of Proposition 1 It follows from the two lemmas above. □

Proof of Theorem 1 From the spline approximation theorem in Page 154 in de Boor (2001) and Theorem 5.1 in Huang (2003), we have supremum order of bias $\|\tilde{m}_2(x) - m(x)\|_\infty = O_p(h^2)$. Hence the bias order is negligible compared to the noise order since $h^2 \ll (nh)^{-1/2}\log^{1/2}(N+1)$. Applying Proposition 1

$$\begin{aligned}
&\liminf_{n\to\infty} P\left[m\left(x\right)\in\hat{m}_2\left(x\right)\pm 2\sigma_{n,2}\left(x\right)\left\{\log\left(N+1\right)\right\}^{1/2} d_n\left(\alpha/2\right), \forall x\in\left[0,1\right]\right] \\
&= \liminf_{n\to\infty} P\left[\sup_{x\in[0,1]}\sigma_{n,2}^{-1}\left(x\right)\left|\tilde{\varepsilon}_2\left(x\right)+\tilde{m}_2\left(x\right)-m\left(x\right)\right| \le 2\left\{\log\left(N+1\right)\right\}^{1/2} d_n\left(\frac{\alpha}{2}\right)\right] \\
&= \liminf_{n\to\infty} P\left[\sup_{x\in[0,1]}\left|\sigma_{n,2}^{-1}\left(x\right)\tilde{\varepsilon}_2\left(x\right)\right| \le 2\left\{\log\left(N+1\right)\right\}^{1/2} d_n\left(\alpha/2\right)\right] \ge 1-\alpha.
\end{aligned}$$

□

Acknowledgments The author would like to thank the associate editor and two referees for their constructive comments which helped the author improve the quality of the paper significantly.

References

Beran, J., Feng, Y. (2002a). Local polynomial fitting with long-memory, short-memory and antipersistent errors. *The Annals of the Institute of Statistical Mathematics, 54*, 291–311.

Beran, J., Feng, Y. (2002b). SEMIFAR models—A semiparametric framework for modelling trends, long-range dependence and nonstationarity. *Computational Statistics and Data Analysis, 40*, 393–419.

Bickel, P. J., Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. *Annals of Statistics, 1*, 1071–1095.

Bosq, D. (1996). *Nonparametric statistics for stochastic processes.* New York: Springer.

Cai, Z. (2002). Regression quantiles for time series. *Econometric Theory, 18*, 169–192.

Claeskens, G., Van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics, 31*, 1852–1884.

de Boor, C. (2001). *A practical guide to splines*. New York: Springer.

Diack, C. (2001). Testing the shape of a regression curve. *Comptes Rendus de l'Académie des Sciences. Série I. Mathématique, 333*(7), 677–680.

Fan, J., Gijbels, I. (1996). *Local polynomial modelling and its applications*. London: Chapman and Hall.

Fan, J., Yao, Q. (2003). *Nonlinear time series*. New York: Springer.

Feng, Y. (2004). Simultaneously modeling conditional heteroskedasticity and scale change. *Econometric Theory, 20*, 563–596.

Gantmacher, F. R., Krein, M. G. (1960). *Oszillationsmatrizen, Oszillationskerne und kleine Schwingungen mechanischer Systeme*. Berlin: Akademie.

Härdle, W. (1989). Asymptotic maximal deviation of M-smoothers. *Journal of Multivariate Analysis, 29*, 163–179.

Härdle, W. (1990). *Applied nonparametric regression*. Cambridge: Cambridge University Press.

Härdle, W., Marron, J. S., Yang, L. (1997). Discussion of "Polynomial splines and their tensor products in extended linear modeling" by Stone et al. *The Annals of Statistics, 25*, 1443–1450.

Härdle, W., Huet, S., Mammen, E., Sperlich, S. (2004). Bootstrap inference in semiparametric generalized additive models. *Econometric Theory, 20*, 265–300.

Huang, J. Z. (1998). Projection estimation in multiple regression with application to functional ANOVA models. *Annals of Statistics, 26*, 242–272.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Annals of Statistics, 31*, 1600–1635.

Huang, J. Z., Yang, L. (2004). Identification of nonlinear additive autoregressive models. *Journal of the Royal Statistical Society Series B, 66*, 463–477.

Huber-Carol, C., Balakrishnan, N., Nikulin, M., Mesbah, M. (2002). *Goodness-of-fit tests and model validity*. Boston: Birkhäuser.

Johnson, R. A., Wichern, D. W. (1992). *Applied multivariate statistical analysis*. New Jersey: Prentice-Hall.

Leadbetter, M. R., Lindgren, G., Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. New York: Springer.

Liang, H., Uña-Álvarez, J. (2009). A Berry–Esseen type bound in kernel density estimation for strong mixing censored samples. *Journal of Multivariate Analysis, 100*, 1219–1231.

Liebscher, E. (1999). Asymptotic normality of nonparametric estimators under mixing condition. *Statistics & Probability Letters, 43*, 243–250.

Liebscher, E. (2001). Estimation of the density and the regression function under mixing conditions. *Statistics & Decisions 19*, 9–26.

Masry, E., Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics, 24*, 165–179.

Olson, J. M., Alagarswamy, G., Andresen, J. A., Campbell, D. J., Davis, A. Y., Ge, J., et al. (2008). Integrating diverse methods to understand climate–land interactions in East Africa. *Geoforum, 39*, 898–911.

Paparoditis, E., Politis, D. (2000). The local bootstrap for kernel estimators under general dependence conditions. *The Annals of Institute of Statistical Mathematics, 52*, 139–259.

Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random Variables. *Journal of Statistical Planning and Inference, 18*, 135–149.

Roussas, G. G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Processes and Their Applications, 36*, 107–116.

Schumway R., Stoffer D. (2006). *Time series analysis and its applications*. New York: Springer.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Song, Q., Yang, L. (2009). Spline confidence bands for variance function. *Journal of Nonparametric Statistics, 21*, 589–609.

Stone, C. J. (1985). Additive regression and other nonparametric models. *Annals of Statistics, 13*, 689–705.

Stone, C. J. (1994). The use of polynomial splines and their tensor products in multivariate function estimation. *Annals of Statistics, 22*, 118–184.

Sunklodas, J. (1984). On the rate of convergence in the central limit theorem for strongly mixing random variables. *Lithuanian Mathematical Journal, 24*, 182–190.

Tusnády, G. (1977). A remark on the approximation of the sample df in the multidimensional case. *Periodica Mathematica Hungarica, 8*, 53–55.

Wang, J. (2009). Modelling time trend via spline confidence band. Manuscript, 26 pages. http://www.math.uic.edu/~wangjing/bandfixedfull.pdf.

Wang, J., Yang, L. (2009a). Polynomial spline confidence bands for regression curves. *Statistica Sinica, 19*, 325–342.

Wang, J., Yang, L. (2009b). Efficient and fast spline-backfitted kernel smoothing of additive models. *Annals of the Institute of Statistical Mathematics, 61*, 663–690.

Wang, J., Qi, J., Yang, L., Olson, J., Nathan, M., Nathan, T., et al. (2006). Derivation of phenological information from remotely sensed imagery for improved regional climate modeling. Manuscript.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society: Series B, 60*, 797–811.

Xue, L., Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica, 16*, 1423–1446.

Yang, L. (2008). Confidence band for additive regression model. *Journal of Data Science, 6*, 207–217.

Zhang, F. (1999). *Matrix theory: Basic results and techniques*. New York: Springer.

Zhou, S., Shen, X., Wolfe, D. A. (1998). Local asymptotics of regression splines and confidence regions. *The Annals of Statistics, 26*, 1760–1782.